**Supplementary Materials**

# Assessment of template based protein structure predictions in CASP10

Yuanpeng J. Huang**, Binchen Mao**, James M. Aramini,

and Gaetano T Montelione

---

Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey and Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, and Northeast Structural Genomics Consortium, 679 Hoes Lane, Piscataway, New Jersey, 08854, USA

** YJH and BM contributed equally to this study and are designated as co-first authors.

## Table of Contents

## Evaluation of RPF Scores

*Selection of $D_{max}$*

In computing RPF scores for CASP10, we selected $D_{max}$ based on its sensitivity in distinguishing the accuracy of the overall structure. Supplementary Figure S1 shows the RPF scores for CASP9-T0570 target with $D_{max}$ ranging from 5 Å to 20 Å. $D_{max}$ = 9.0 Å gives the highest discrimination for correct core fold, excluding the loop regions. For example, at 5.0 Å cutoff, the RPF DP scores for models 264_1 and 250_1 are very close, indicating that 5.0 Å mainly measures local information, while longer range of distances are needed for core fold comparisons. The 9.0 Å cutoff gives the largest differences in DP scores among the models with GDT_TS scores > 80 (e.g. 361_1, 481_1, 276_1), > 50 (e.g. 250_1) and < 40 groups (e.g. 264_1). We also examined a set of CASD-NMR models [1] with different cutoff ranges and similar sensitivities were observed at the 9.0 Å cutoff. Ideally, we would like to find a cutoff, which can not only distinguish good models from bad models, but also find better models among the good models.

Supplementary Figure S1 shows that at distance cutoff of 5.0, RPF is dominated by the local side-chain atom pair packing information. As the distance cutoff increases, the correctness of fold starts to contribute to the RPF score. We choose a distance cutoff of 9.0 Å, which seems to be a good balance of both global fold (main chain conformation) and also local side-chain information. For difficult targets with poor quality, the RPF score will be dominated by the main chain conformation, not the local side-chain packing. In comparing models which all have similar overall fold accuracy, the RPF score will give higher scores for models with better local side-chain packing.

A similar distance network based method, LDDT, was used in CASP9[2]. The two major differences between RPF and LDDT methods are: (1) the LDDT is similar to the Recall measure of RPF, but LDDT uses exact distance comparisons (i.e. distance difference thresholds of 0.5, 1.0, 2.0 and 4.0 Å[2]), while RPF identifies distances as TPs if they are within a distance upper bound. (2) the LDDT score used a cutoff of < 5 Å in CASP9, which is more focused on local distance comparisons. For example, the CASP9 LDDT-5 scores are 81.0 and 66.8 for targets 481_1 and 276_1, respectively, even though target 276_1 has a slightly higher GDT-TS score than 481_1. The RPF DP score for 276_1 is slightly smaller than 481_1 at $D_{max}$ = 9 Å (Figure. S1). Even at $D_{max}$ = 5.0 Å, the differences of RPF scores for 276_1 and 481_1 (0.6 and 0.64, respectively) are not as large as the differences of the CASP9 LDDT-5 scores (66.8 and 81.0, respectively), reflecting the generally high weight on local structure by LDDT-5. For CASP10 assessment, the LDDT cutoff was increased to 15 Å in the result provided by the CASP Prediction Center.

*Comparison with GDT based methods*

The correlations among the DP score and GDT based methods are very strong. The Spearman's correlation coefficient is 0.93 between the GDT_HA and DP scores. However, as illustrated in Supplemental Figure S4, for some structural differences (e.g. core packing) the DP score is more sensitive, while for other (e.g. helix tilt angles) the GDT-HA score is more sensitive.

*Loops* – Flexible loop regions tend to have fewer close distances than core regions. Models with similar cores, but which differ in the loop regions may have similar RPF

scores. However, the GDT based measure will give a higher score for the model with loop conformation similar to the target structure. As an example, Supplemental Figure S4A shows two models with very similar core packing, but with very different loop structures, especially the N- and C-terminal loops. Model 113_1(green) has GDT_HA = 67.73 and RPF=0.74. Model 101_1(cyan) has GDT_HA = 64.19 and RPF=0.81. Model 113_1 is the top pick by GDT_HA, which has the best alignment at the N-terminal loops. RPF, on the other hand, identifies 101_1 as the top model, as it has more accurate core packing, although its N-terminal loop structure is quite different from the target structure.

*Helix Tilt Angles* - Distance network based measures will be less sensitive to the helix tilt angle than the GDT based measure. An example is shown in Figure S4B for AU target T0678-D1. Two models 079_1 and 237_1 have very similar fold. The helix cores of the two models are similar to the target structure, except that the helices have significantly different tilt angles. For both models, The GDT_HA scores are low (i.e. 24.35 and 22.73 for 079_1 and 237_1), while the RPF-9 scores are higher (i.e. 0.56 and 0.55, respectively), largely because the distance networks within the cores of helix packing are both similar to that of the target structure. However, the differences in RPF-9 scores between prediction models 079_1 and 237_1 are relatively small because the main differences involve small differences in helix tilt angles, which do not significantly affect the RPF DP scores.

*Normalization against random structures*

Unlike other metrics used in CASP, the RPF score is normalized against a free rotating chain model. For this reason, RPF DP scores are very discriminative against

random structures. Random-like models will have RPF scores very close to zero. Structures with incorrect secondary structures or incorrect folds can even have negative RPF scores, indicating that they are even worse than random structures. For example, CASP10 group 027 models for targets T0671 and T0705 contain mostly random loops (Supplemental Figure S3, right). The GDT_HA scores are 10.23 and 11.98 accordingly. The core structures of CASP10 group 237 for targets T0671-D1 and T0705-D1 are much similar to the target structures (Supplemental Figure S3, left) than CASP10 group 027_1 models. The GDT_HA scores for 237_1 are 28.7 and 32.81 for T0671 and T0705, which is only about 3 fold higher than the scores for 027_1. The RPF-9 scores for T0671 are 0.57 for group 237 and 0.08 for 027_1; the RPF-9 scores for T0705 are 0.49 for 237_1 and 0.09 for 027_1. The RPF-9 scores of 237_1 are about 5 fold higher than the RPF-9 scores of 027_1, demonstrating that the RPF-9 scores normalized to random structure (i.e. the RPF-9 DP scores) have stronger discriminating power than the GDT_HA scores against structures with random-like incorrect folds.

The scatter plot of RPF raw scores (Supplemental Figure S5) for the 57 hAUs between the zhang (237) and LeeCon (027) groups suggests that the biggest differences in the RPF-9 DP scores are for targets T0671-D1 and T0705-D1 (discussed above), which influences the relative ranking of these two groups (Figure 4B and Table II) in the Human and/or Server group.

Supplementary Table S1. Sum and Average Z-scores for All Server Predictor Groups – 112 hsAU Targets.

| Name | N_model | GDT-HA | GDC-all | RPF | LDDT | Sum | Avg_4a | Avg_4s | MolPro |
|---|---|---|---|---|---|---|---|---|---|
| BAKER-ROSETTASERVER | 112 | 52.03 | 62.49 | 77.60 | 75.18 | 66.83 | 0.60 | 0.60 | 219.08 |
| Zhang-Server | 112 | 54.21 | 48.41 | 78.52 | 66.59 | 61.93 | 0.55 | 0.55 | 15.62 |
| PMS | 112 | 37.21 | 49.23 | 72.74 | 74.08 | 58.32 | 0.52 | 0.52 | 2.03 |
| QUARK | 111 | 50.68 | 41.91 | 70.40 | 60.60 | 55.90 | 0.50 | 0.50 | 9.40 |
| HHpred-thread | 111 | 44.39 | 48.89 | 50.60 | 58.05 | 50.48 | 0.45 | 0.46 | -144.25 |
| RaptorX-ZY | 112 | 43.44 | 42.14 | 53.96 | 47.22 | 46.69 | 0.42 | 0.42 | -70.08 |
| HHpredA | 112 | 44.20 | 46.70 | 45.30 | 49.21 | 46.35 | 0.41 | 0.41 | -137.90 |
| HHpredAQ | 112 | 40.82 | 43.87 | 44.73 | 48.88 | 44.58 | 0.40 | 0.40 | -139.75 |
| RaptorX | 112 | 42.91 | 43.50 | 42.62 | 36.23 | 41.32 | 0.37 | 0.37 | -36.43 |
| MULTICOM-NOVEL | 112 | 31.30 | 34.88 | 45.11 | 47.34 | 39.66 | 0.35 | 0.35 | 27.56 |
| MULTICOM-REFINE | 112 | 27.70 | 33.03 | 39.74 | 42.52 | 35.75 | 0.32 | 0.32 | 30.97 |
| MULTICOM-CLUSTER | 112 | 24.39 | 30.72 | 39.49 | 41.41 | 34.00 | 0.30 | 0.30 | 30.46 |
| TASSER-VMT | 112 | 29.13 | 24.93 | 48.74 | 28.75 | 32.89 | 0.29 | 0.29 | -88.38 |
| PconsM | 112 | 24.21 | 25.59 | 43.61 | 37.44 | 32.71 | 0.29 | 0.29 | 27.14 |
| chunk-TASSER | 112 | 24.85 | 29.28 | 38.73 | 36.42 | 32.32 | 0.29 | 0.29 | -28.19 |
| Pcons-net | 112 | 13.93 | 20.25 | 32.86 | 28.66 | 23.93 | 0.21 | 0.21 | 65.23 |
| Mufold-MD | 112 | 10.72 | 16.35 | 32.47 | 30.63 | 22.54 | 0.20 | 0.20 | 93.95 |
| MULTICOM-CONSTRUCT | 112 | 14.40 | 19.42 | 26.07 | 29.11 | 22.25 | 0.20 | 0.20 | 16.48 |
| MUFOLD-Server | 112 | 16.05 | 15.96 | 27.70 | 23.79 | 20.88 | 0.19 | 0.19 | -13.69 |
| Seok-server | 112 | 13.31 | 23.50 | 12.71 | 25.40 | 18.73 | 0.17 | 0.17 | 75.22 |
| FALCON-TOPO | 112 | 6.67 | 7.67 | 17.54 | 11.72 | 10.90 | 0.10 | 0.10 | -30.77 |
| FALCON-TOPO-X | 112 | 3.35 | 5.14 | 11.72 | 5.90 | 6.53 | 0.06 | 0.06 | -36.42 |
| PconsD | 111 | 0.89 | -0.42 | 14.28 | 8.15 | 5.73 | 0.05 | 0.05 | 0.36 |
| Phyre2_A | 112 | 7.88 | 5.33 | 2.20 | 6.29 | 5.43 | 0.05 | 0.05 | -82.57 |
| ZHOU-SPARKS-X | 112 | -3.20 | -5.76 | 11.41 | 8.48 | 2.73 | 0.02 | 0.02 | -68.11 |
| YASARA | 73 | -7.06 | 1.75 | 4.10 | 9.00 | 1.95 | 0.02 | 0.03 | 166.40 |
| SAM-T08-server | 101 | -13.52 | -3.94 | 7.48 | 7.69 | -0.57 | -0.01 | -0.01 | 19.92 |
| RaptorX-Roll | 12 | -5.06 | -5.29 | -3.29 | -3.36 | -4.25 | -0.04 | -0.35 | -1.31 |
| IntFOLD2 | 112 | -9.66 | -8.22 | -2.28 | -7.93 | -7.02 | -0.06 | -0.06 | -48.04 |
| Bhageerath_abinitio | 5 | -8.13 | -9.67 | -9.47 | -8.56 | -8.96 | -0.08 | -1.79 | 3.32 |
| FFAS03mt | 101 | 1.41 | 1.58 | -19.01 | -20.22 | -9.06 | -0.08 | -0.09 | -25.56 |
| sysimm | 56 | -3.17 | -1.12 | -17.31 | -17.92 | -9.88 | -0.09 | -0.18 | 90.96 |
| Distill | 112 | -3.66 | -8.55 | -15.15 | -13.34 | -10.18 | -0.09 | -0.09 | 9.13 |
| slbio | 104 | -12.25 | -7.95 | -9.27 | -17.04 | -11.63 | -0.10 | -0.11 | 10.52 |
| chuo-fams-server | 111 | -17.29 | -11.39 | -14.24 | -14.29 | -14.30 | -0.13 | -0.13 | -73.26 |
| MATRIX | 102 | -19.39 | -14.58 | -22.75 | -18.70 | -18.86 | -0.17 | -0.19 | 42.65 |
| FFAS03c | 111 | -14.74 | -13.38 | -22.92 | -25.09 | -19.03 | -0.17 | -0.17 | -57.93 |
| samcha-server | 101 | -24.72 | -22.03 | -12.33 | -18.00 | -19.27 | -0.17 | -0.19 | -45.81 |
| Bilab-ENABLE | 112 | -26.80 | -21.76 | -24.68 | -20.42 | -23.42 | -0.21 | -0.21 | 21.63 |
| GSmetaserver | 70 | -15.33 | -14.73 | -28.94 | -35.01 | -23.50 | -0.21 | -0.34 | -40.75 |
| hGen3D | 112 | -9.47 | -30.29 | -28.91 | -32.41 | -25.27 | -0.23 | -0.23 | -58.42 |
| chuo-repack-server | 112 | -29.23 | -21.85 | -29.90 | -30.53 | -27.88 | -0.25 | -0.25 | -81.85 |
| NewSerf | 112 | -11.62 | -28.62 | -35.97 | -35.32 | -27.88 | -0.25 | -0.25 | -75.67 |
| IntFOLD | 112 | -23.39 | -27.32 | -30.51 | -32.80 | -28.51 | -0.26 | -0.26 | -27.94 |
| Distill_roll | 112 | -37.04 | -34.25 | -24.07 | -32.30 | -31.92 | -0.29 | -0.29 | 32.17 |
| 3D-JIGSAW_V5-0 | 108 | -29.05 | -22.28 | -39.54 | -40.05 | -32.73 | -0.29 | -0.30 | -16.98 |
| Atome2_CBS | 101 | -28.93 | -29.74 | -35.04 | -41.75 | -33.87 | -0.30 | -0.34 | -59.41 |
| FRESS_server | 111 | -57.01 | -49.66 | -8.83 | -22.24 | -34.44 | -0.31 | -0.31 | -156.24 |
| FFAS03hj | 103 | -28.52 | -28.11 | -36.56 | -45.10 | -34.57 | -0.31 | -0.34 | -32.15 |
| FFAS03 | 102 | -23.52 | -20.25 | -44.76 | -50.20 | -34.68 | -0.31 | -0.34 | -35.48 |
| STRINGS | 99 | -44.64 | -44.96 | -23.90 | -33.19 | -36.67 | -0.33 | -0.37 | -10.98 |
| MUFold_CRF | 109 | -35.22 | -36.29 | -37.87 | -40.72 | -37.53 | -0.34 | -0.34 | -54.80 |
| UGACSBL | 102 | -34.39 | -40.31 | -37.69 | -39.48 | -37.97 | -0.34 | -0.37 | -43.39 |
| AOBA-server | 110 | -36.94 | -38.38 | -35.16 | -43.72 | -38.55 | -0.34 | -0.35 | -46.23 |
| panther | 85 | -40.25 | -39.33 | -58.91 | -56.60 | -48.77 | -0.44 | -0.57 | -58.87 |
| PROTAGORAS | 94 | -50.77 | -52.06 | -54.20 | -51.59 | -52.16 | -0.47 | -0.56 | -53.52 |
| Jiang_Fold | 112 | -56.19 | -57.52 | -47.08 | -49.90 | -52.67 | -0.47 | -0.47 | -51.49 |
| SAM-T06-server | 101 | -65.56 | -58.48 | -60.48 | -52.06 | -59.15 | -0.53 | -0.59 | 32.25 |
| Jiang_Server | 112 | -50.44 | -65.56 | -66.89 | -67.40 | -62.57 | -0.56 | -0.56 | -57.28 |
| Lenserver | 38 | -61.26 | -62.42 | -70.57 | -74.78 | -67.26 | -0.60 | -1.77 | -76.00 |
| Jiang_Threader | 112 | -56.01 | -66.20 | -83.38 | -78.82 | -71.10 | -0.64 | -0.64 | -58.21 |
| confuzz3d | 41 | -70.60 | -69.72 | -72.43 | -71.81 | -71.14 | -0.64 | -1.74 | -70.39 |
| BhageerathH | 111 | -87.57 | -83.65 | -107.88 | -94.63 | -93.43 | -0.83 | -0.84 | 92.88 |
| confuzzGS | 55 | -96.87 | -95.47 | -97.32 | -96.74 | -96.60 | -0.86 | -1.76 | -75.50 |
| HOMER | 84 | -76.92 | -85.26 | -122.78 | -125.69 | -102.66 | -0.92 | -1.22 | 59.41 |
| RBO-MBS | 107 | -150.06 | -148.38 | -144.83 | -147.66 | -147.73 | -1.32 | -1.38 | 242.75 |
| RBO-i-MBS | 108 | -151.39 | -149.17 | -146.24 | -147.41 | -148.55 | -1.33 | -1.38 | 242.35 |
| RBO-MBS-BB | 108 | -153.63 | -153.27 | -150.25 | -150.53 | -151.92 | -1.36 | -1.41 | 242.55 |
| RBO-i-MBS-BB | 108 | -159.76 | -157.54 | -155.52 | -156.68 | -157.38 | -1.41 | -1.46 | 240.79 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 112 hsAUs.

Supplementary Table S2. Sum and Average Z-scores for All Predictor Groups – 57 hAU Targets.

| Group | Name | N_model | GDT-HA | GDC-all | RPF | LDDT | Sum | Avg_4a | Avg_4s | MolPro |
|---|---|---|---|---|---|---|---|---|---|---|
| 237 | zhang | 57 | 37.64 | 33.50 | 50.23 | 46.15 | 41.88 | 0.74 | 0.74 | 1.63 |
| 27 | LEEcon | 57 | 35.60 | 36.14 | 45.39 | 47.43 | 41.14 | 0.72 | 0.72 | 12.84 |
| 035s | Zhang-Server | 57 | 32.85 | 31.00 | 44.68 | 40.19 | 37.18 | 0.65 | 0.65 | -2.33 |
| 130 | Pcomb | 57 | 29.92 | 28.99 | 44.96 | 41.68 | 36.39 | 0.64 | 0.64 | 22.33 |
| 197 | Mufold | 56 | 30.46 | 29.74 | 42.41 | 39.20 | 35.45 | 0.62 | 0.63 | -9.75 |
| 79 | TASSER | 57 | 34.34 | 32.49 | 40.59 | 33.92 | 35.34 | 0.62 | 0.62 | -32.59 |
| 267 | Pcons | 56 | 28.76 | 28.78 | 42.48 | 40.73 | 35.19 | 0.62 | 0.63 | 15.26 |
| 489 | MULTICOM | 57 | 29.85 | 29.53 | 40.17 | 39.02 | 34.64 | 0.61 | 0.61 | 9.01 |
| 344 | Jones-UCL | 56 | 26.78 | 28.72 | 43.08 | 37.82 | 34.10 | 0.60 | 0.61 | -50.10 |
| 114s | QUARK | 56 | 29.68 | 25.98 | 38.50 | 35.87 | 32.51 | 0.57 | 0.58 | -5.65 |
| 301 | LEE | 57 | 27.53 | 29.38 | 34.98 | 37.52 | 32.35 | 0.57 | 0.57 | 6.37 |
| 477 | BAKER | 57 | 31.46 | 29.24 | 35.34 | 33.04 | 32.27 | 0.57 | 0.57 | 104.97 |
| 475 | CNIO | 57 | 27.86 | 23.52 | 39.78 | 37.69 | 32.21 | 0.57 | 0.57 | -2.34 |
| 350 | Kloczkowski_Lab | 57 | 26.74 | 23.82 | 40.01 | 36.94 | 31.88 | 0.56 | 0.56 | 18.79 |
| 490 | Zhang_Refinement | 57 | 26.03 | 24.81 | 33.25 | 32.61 | 29.18 | 0.51 | 0.51 | 16.67 |
| 294 | chuo-repack | 57 | 23.19 | 20.55 | 36.79 | 34.19 | 28.68 | 0.50 | 0.50 | -6.90 |
| 458 | Sternberg | 57 | 25.94 | 23.30 | 32.25 | 31.66 | 28.29 | 0.50 | 0.50 | -10.66 |
| 365 | chuo-fams | 57 | 22.50 | 21.08 | 35.92 | 31.92 | 27.86 | 0.49 | 0.49 | 4.58 |
| 428 | PconsQ | 56 | 22.08 | 20.09 | 33.92 | 33.43 | 27.38 | 0.48 | 0.49 | 7.49 |
| 434 | chuo-fams-consensus | 57 | 20.19 | 15.16 | 32.68 | 31.41 | 24.86 | 0.44 | 0.44 | -7.49 |
| 481 | Chicken_George | 57 | 21.70 | 18.50 | 31.02 | 28.17 | 24.85 | 0.44 | 0.44 | 5.58 |
| 122s | RaptorX-ZY | 57 | 25.88 | 24.34 | 24.68 | 21.98 | 24.22 | 0.43 | 0.43 | -44.38 |
| 45 | Zhang_Ab_Initio | 57 | 21.07 | 20.37 | 28.28 | 26.96 | 24.17 | 0.42 | 0.42 | 4.98 |
| 285 | McGuffin | 55 | 18.28 | 16.26 | 30.52 | 27.92 | 23.25 | 0.41 | 0.42 | -1.60 |
| 405 | Mufold2 | 54 | 18.98 | 18.36 | 28.92 | 25.93 | 23.05 | 0.40 | 0.43 | -17.03 |
| 330s | BAKER-ROSETTASERVER | 57 | 18.90 | 20.60 | 27.45 | 25.02 | 22.99 | 0.40 | 0.40 | 106.59 |
| 108s | PMS | 57 | 17.29 | 19.08 | 26.26 | 28.70 | 22.83 | 0.40 | 0.40 | -1.68 |
| 26 | ProQ2clust | 56 | 18.95 | 20.74 | 22.97 | 23.47 | 21.53 | 0.38 | 0.38 | -7.31 |
| 315 | keasar | 54 | 13.37 | 15.59 | 31.75 | 25.28 | 21.50 | 0.38 | 0.40 | -57.45 |
| 388 | ProQ2 | 57 | 13.29 | 13.24 | 29.50 | 25.80 | 20.46 | 0.36 | 0.36 | 49.01 |
| 486s | RaptorX | 57 | 25.59 | 26.28 | 15.40 | 14.06 | 20.33 | 0.36 | 0.36 | -22.69 |
| 317 | SHORTLE | 52 | 18.71 | 14.31 | 24.49 | 21.13 | 19.66 | 0.35 | 0.38 | 26.65 |
| 101 | WeFold | 52 | 14.58 | 12.55 | 27.81 | 20.54 | 18.87 | 0.33 | 0.36 | -37.91 |
| 164 | 4_BODY_POTENTIALS | 55 | 13.16 | 12.88 | 29.01 | 20.13 | 18.80 | 0.33 | 0.34 | 5.89 |
| 370s | HHpred-thread | 56 | 18.49 | 18.23 | 15.33 | 19.83 | 17.97 | 0.32 | 0.32 | -85.84 |
| 493 | LEEMO | 57 | 15.71 | 13.09 | 19.45 | 21.19 | 17.36 | 0.31 | 0.31 | -21.57 |
| 281 | Ariadne | 56 | 13.41 | 13.35 | 20.22 | 21.26 | 17.06 | 0.30 | 0.31 | 19.05 |
| 335s | TASSER-VMT | 57 | 17.23 | 16.50 | 19.15 | 12.73 | 16.40 | 0.29 | 0.29 | -42.93 |
| 430s | HHpredA | 57 | 16.03 | 17.80 | 10.79 | 15.19 | 14.95 | 0.26 | 0.26 | -80.89 |
| 223s | HHpredAQ | 57 | 15.82 | 17.58 | 10.78 | 15.17 | 14.84 | 0.26 | 0.26 | -80.79 |
| 103s | PconsM | 57 | 10.05 | 10.20 | 16.48 | 14.44 | 12.79 | 0.22 | 0.22 | 16.95 |
| 424s | MULTICOM-NOVEL | 57 | 10.16 | 10.17 | 14.64 | 15.57 | 12.64 | 0.22 | 0.22 | 10.21 |
| 473 | Seok | 56 | 13.11 | 14.88 | 8.69 | 13.30 | 12.50 | 0.22 | 0.22 | 38.77 |
| 280 | ProQ2clust2 | 57 | 9.40 | 8.38 | 15.39 | 14.82 | 12.00 | 0.21 | 0.21 | -9.88 |
| 125s | MULTICOM-REFINE | 57 | 8.31 | 8.90 | 12.72 | 14.14 | 11.02 | 0.19 | 0.19 | 9.88 |
| 081s | MULTICOM-CLUSTER | 57 | 8.41 | 9.35 | 12.59 | 13.20 | 10.89 | 0.19 | 0.19 | 12.45 |
| 479 | Boniecki_LoCoGRef | 41 | 5.90 | 7.37 | 14.77 | 13.89 | 10.48 | 0.18 | 0.26 | -17.45 |
| 488s | chunk-TASSER | 57 | 7.73 | 8.22 | 10.10 | 9.37 | 8.86 | 0.16 | 0.16 | -21.38 |
| 292s | Pcons-net | 57 | 5.21 | 7.02 | 12.15 | 9.61 | 8.50 | 0.15 | 0.15 | 39.17 |
| 435 | ossia | 57 | 7.25 | 7.48 | 8.65 | 9.07 | 8.11 | 0.14 | 0.14 | -18.63 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 57 hAUs.

Supplementary Table S2. Sum and Average Z-scores for All Predictor Groups – 57 hAU Targets.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 222s | MULTICOM-CONSTRUCT | 57 | 4.32 | 5.91 | 8.39 | 11.32 | 7.49 | 0.13 | 0.13 | 3.68 |
| 113s | SAM-T08-server | 47 | 4.29 | 6.28 | 8.60 | 7.58 | 6.69 | 0.12 | 0.14 | 9.12 |
| 178 | Bilab | 52 | 2.67 | 3.46 | 6.78 | 9.34 | 5.56 | 0.10 | 0.11 | 2.16 |
| 68 | FOLDIT | 10 | 3.94 | 4.61 | 4.86 | 4.44 | 4.46 | 0.08 | 0.45 | 16.74 |
| 251 | laufercenter_meta | 8 | 3.23 | 4.30 | 3.43 | 3.61 | 3.64 | 0.06 | 0.46 | 17.72 |
| 165 | Void_Crushers | 10 | 1.59 | 0.68 | 4.64 | 3.73 | 2.66 | 0.05 | 0.27 | 18.50 |
| 261s | Seok-server | 57 | 5.61 | 8.93 | -5.62 | 1.80 | 2.68 | 0.05 | 0.05 | 32.53 |
| 141 | Bates_BMM | 55 | 1.66 | -0.51 | 5.94 | 2.80 | 2.47 | 0.04 | 0.05 | -23.97 |
| 333s | MUFOLD-Server | 57 | 2.04 | 2.45 | 1.09 | 0.98 | 1.64 | 0.03 | 0.03 | -9.28 |
| 437 | sumalab | 57 | 4.86 | 4.16 | -1.79 | -2.92 | 1.08 | 0.02 | 0.02 | -1.82 |
| 286s | Mufold-MD | 57 | -0.64 | 0.63 | 2.21 | 1.74 | 0.99 | 0.02 | 0.02 | 36.21 |
| 290 | Wallner | 1 | 0.73 | 0.93 | 1.17 | 0.99 | 0.96 | 0.02 | 0.96 | 1.86 |
| 453 | KnowMIN | 57 | -1.22 | -1.05 | 4.52 | 1.01 | 0.82 | 0.01 | 0.01 | 41.12 |
| 260 | wfFUGT | 9 | 0.80 | -0.79 | 1.01 | 1.51 | 0.63 | 0.01 | 0.07 | -6.11 |
| 341 | Contenders | 9 | -1.07 | -0.68 | 1.93 | 1.12 | 0.33 | 0.01 | 0.04 | 16.47 |
| 30 | TAUbioinfounit | 3 | -0.16 | -0.35 | -0.15 | 0.07 | -0.15 | 0.00 | -0.05 | 5.66 |
| 10 | TSlab-refine | 1 | -0.31 | -0.25 | -0.34 | -0.45 | -0.34 | -0.01 | -0.34 | 1.23 |
| 85 | Anthropic_Dreams | 9 | -2.00 | -2.42 | 1.83 | 0.65 | -0.49 | -0.01 | -0.05 | 15.97 |
| 348s | Phyre2_A | 57 | 1.43 | 1.23 | -5.20 | 0.08 | -0.62 | -0.01 | -0.01 | -45.05 |
| 29 | shisen | 43 | -3.60 | -1.81 | 0.78 | 1.34 | -0.82 | -0.01 | -0.02 | 47.84 |
| 298 | MidwayFolding | 50 | 12.86 | 7.30 | -8.86 | -14.74 | -0.86 | -0.02 | -0.02 | -47.11 |
| 149 | wfFUIK | 10 | -1.86 | -2.47 | -1.17 | -1.28 | -1.70 | -0.03 | -0.17 | 11.59 |
| 215 | ppfld | 9 | -1.95 | -0.90 | -2.27 | -2.43 | -1.89 | -0.03 | -0.21 | -0.24 |
| 048s | Bhageerath_abinitio | 2 | -2.19 | -3.36 | -3.54 | -2.62 | -2.93 | -0.05 | -1.46 | 0.62 |
| 77 | FLOUDAS | 41 | -6.44 | -7.35 | 0.47 | -2.64 | -3.99 | -0.07 | -0.10 | 26.77 |
| 411s | FALCON-TOPO | 57 | -5.05 | -4.48 | -1.89 | -5.30 | -4.18 | -0.07 | -0.07 | -16.54 |
| 444 | Lenregular | 3 | -3.48 | -4.06 | -4.79 | -6.00 | -4.58 | -0.08 | -1.53 | -6.00 |
| 356s | sysimm | 25 | -0.64 | 2.25 | -10.33 | -12.51 | -5.31 | -0.09 | -0.21 | 45.02 |
| 124s | PconsD | 56 | -9.02 | -9.24 | 0.25 | -4.32 | -5.58 | -0.10 | -0.10 | -4.48 |
| 358s | RaptorX-Roll | 12 | -6.15 | -5.92 | -5.25 | -5.15 | -5.62 | -0.10 | -0.47 | -2.32 |
| 441 | WeFoldMix | 5 | -6.16 | -5.95 | -5.04 | -5.44 | -5.65 | -0.10 | -1.13 | 8.72 |
| 131 | BioNanopore | 10 | -5.85 | -5.14 | -5.76 | -6.92 | -5.92 | -0.10 | -0.59 | -2.70 |
| 311 | Laufer | 9 | -6.47 | -7.60 | -5.69 | -4.33 | -6.02 | -0.11 | -0.67 | 8.05 |
| 413s | ZHOU-SPARKS-X | 57 | -9.65 | -10.22 | -2.91 | -2.29 | -6.27 | -0.11 | -0.11 | -38.45 |
| 43 | sessions | 56 | -11.60 | -10.17 | -4.08 | -4.80 | -7.66 | -0.13 | -0.14 | -30.36 |
| 466 | Taylor | 9 | -7.14 | -7.75 | -9.40 | -6.90 | -7.80 | -0.14 | -0.87 | -13.63 |
| 373 | Kim_Kihara | 57 | -12.32 | -12.69 | -3.24 | -5.17 | -8.36 | -0.15 | -0.15 | 56.17 |
| 456s | FALCON-TOPO-X | 57 | -8.92 | -7.97 | -7.64 | -10.80 | -8.83 | -0.16 | -0.16 | -22.11 |
| 448 | KIAS-Gdansk | 10 | -9.86 | -10.30 | -8.58 | -8.31 | -9.26 | -0.16 | -0.93 | -15.49 |
| 028s | YASARA | 23 | -8.78 | -10.62 | -10.06 | -7.62 | -9.27 | -0.16 | -0.40 | 49.58 |
| 072s | Distill | 57 | -6.56 | -6.58 | -11.64 | -12.52 | -9.33 | -0.16 | -0.16 | -3.10 |
| 273s | IntFOLD2 | 57 | -10.23 | -8.01 | -10.10 | -10.23 | -9.64 | -0.17 | -0.17 | -26.47 |
| 51 | thyzju | 8 | -9.83 | -9.27 | -9.44 | -10.33 | -9.72 | -0.17 | -1.22 | -4.63 |
| 494s | GSmetaserver | 30 | -6.29 | -5.85 | -14.29 | -18.40 | -11.21 | -0.20 | -0.37 | -19.70 |
| 155 | Graham | 7 | -11.09 | -9.65 | -11.88 | -12.74 | -11.34 | -0.20 | -1.62 | -13.13 |
| 24 | POEMhome | 35 | -10.20 | -7.82 | -14.02 | -14.02 | -11.52 | -0.20 | -0.33 | -26.38 |
| 302s | 3D-JIGSAW_V5-0 | 53 | -8.30 | -7.22 | -16.44 | -20.33 | -13.07 | -0.23 | -0.25 | -13.14 |
| 175s | FRESS_server | 56 | -20.51 | -17.40 | -10.19 | -5.53 | -13.41 | -0.24 | -0.24 | -83.74 |
| 198s | chuo-fams-server | 56 | -14.56 | -11.40 | -14.98 | -13.72 | -13.67 | -0.24 | -0.24 | -47.41 |
| 172 | Zhang-IRU | 45 | -17.01 | -18.00 | -10.51 | -13.17 | -14.67 | -0.26 | -0.33 | 108.42 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 57 hAUs.

Supplementary Table S2. Sum and Average Z-scores for All Predictor Groups – 57 hAU Targets.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 300 | PAX | 53 | -12.87 | -10.37 | -20.63 | -17.60 | -15.37 | -0.27 | -0.29 | 47.25 |
| 287 | wfCPUNK | 13 | -16.62 | -16.03 | -15.23 | -14.93 | -15.70 | -0.28 | -1.21 | -12.02 |
| 115s | slbio | 54 | -11.69 | -8.43 | -17.16 | -25.63 | -15.73 | -0.28 | -0.29 | 11.35 |
| 221s | Atome2_CBS | 48 | -14.70 | -13.90 | -17.52 | -19.88 | -16.50 | -0.29 | -0.34 | -38.28 |
| 163s | hGen3D | 57 | -16.36 | -19.18 | -17.29 | -18.49 | -17.83 | -0.31 | -0.31 | -32.39 |
| 121s | STRINGS | 44 | -21.39 | -21.11 | -14.54 | -16.18 | -18.31 | -0.32 | -0.42 | -8.83 |
| 112s | samcha-server | 52 | -19.06 | -16.85 | -17.55 | -20.03 | -18.37 | -0.32 | -0.35 | -20.19 |
| 148s | MUFold_CRF | 56 | -15.68 | -14.95 | -21.26 | -21.61 | -18.38 | -0.32 | -0.33 | -30.54 |
| 457s | PROTAGORAS | 40 | -18.29 | -15.92 | -21.01 | -18.24 | -18.37 | -0.32 | -0.46 | -28.27 |
| 343s | NewSerf | 57 | -14.37 | -16.17 | -22.18 | -21.53 | -18.56 | -0.33 | -0.33 | -45.87 |
| 462s | confuzz3d | 14 | -18.52 | -17.26 | -20.20 | -19.92 | -18.98 | -0.33 | -1.36 | -24.80 |
| 464s | UGACSBL | 47 | -20.11 | -21.78 | -23.03 | -20.15 | -21.27 | -0.37 | -0.45 | -21.80 |
| 098s | confuzzGS | 16 | -21.68 | -19.86 | -22.67 | -22.02 | -21.56 | -0.38 | -1.35 | -16.22 |
| 375s | FFAS03c | 56 | -17.74 | -12.33 | -29.19 | -27.81 | -21.77 | -0.38 | -0.39 | -39.46 |
| 439s | MATRIX | 48 | -21.03 | -16.79 | -26.49 | -23.27 | -21.90 | -0.38 | -0.46 | 18.72 |
| 204s | FFAS03mt | 47 | -13.00 | -9.29 | -32.19 | -33.78 | -22.07 | -0.39 | -0.47 | -11.28 |
| 179s | Lenserver | 15 | -17.26 | -17.35 | -26.25 | -29.41 | -22.57 | -0.40 | -1.51 | -30.00 |
| 238s | chuo-repack-server | 57 | -22.43 | -18.09 | -26.61 | -25.49 | -23.16 | -0.41 | -0.41 | -50.14 |
| 277s | Bilab-ENABLE | 57 | -24.69 | -22.73 | -25.74 | -22.62 | -23.95 | -0.42 | -0.42 | 11.13 |
| 275s | FFAS03hj | 49 | -19.59 | -16.96 | -26.59 | -33.46 | -24.15 | -0.42 | -0.49 | -18.91 |
| 087s | Distill_roll | 57 | -30.97 | -28.33 | -16.34 | -24.82 | -25.12 | -0.44 | -0.44 | 10.85 |
| 498s | IntFOLD | 57 | -20.60 | -19.83 | -31.00 | -31.70 | -25.78 | -0.45 | -0.45 | -8.47 |
| 152 | Cornell-Gdansk | 24 | -28.65 | -28.12 | -23.11 | -24.43 | -26.08 | -0.46 | -1.09 | -36.25 |
| 381s | SAM-T06-server | 47 | -27.41 | -23.99 | -29.70 | -26.19 | -26.82 | -0.47 | -0.57 | 16.63 |
| 040s | FFAS03 | 48 | -18.31 | -13.83 | -37.46 | -39.69 | -27.32 | -0.48 | -0.57 | -17.59 |
| 282 | Jiang_Human | 57 | -31.73 | -30.23 | -27.96 | -27.89 | -29.45 | -0.52 | -0.52 | -26.83 |
| 476s | AOBA-server | 55 | -27.96 | -23.90 | -32.24 | -37.97 | -30.52 | -0.54 | -0.56 | -25.30 |
| 246 | MeilerLab | 24 | -31.18 | -30.26 | -34.12 | -31.07 | -31.66 | -0.56 | -1.32 | 10.01 |
| 471 | chuo-binding-sites | 56 | -30.41 | -27.18 | -40.43 | -38.55 | -34.14 | -0.60 | -0.61 | -56.04 |
| 265 | MicroSumaLab | 27 | -33.94 | -31.51 | -36.30 | -35.47 | -34.31 | -0.60 | -1.27 | 27.50 |
| 088s | panther | 37 | -31.01 | -26.74 | -42.31 | -41.93 | -35.50 | -0.62 | -0.96 | -27.53 |
| 482 | biouv | 28 | -35.25 | -33.70 | -39.41 | -44.19 | -38.14 | -0.67 | -1.36 | -48.61 |
| 463s | Jiang_Fold | 57 | -41.49 | -39.41 | -39.10 | -38.30 | -39.58 | -0.69 | -0.69 | -28.72 |
| 116s | HOMER | 38 | -31.62 | -31.80 | -50.51 | -56.24 | -42.54 | -0.75 | -1.12 | 30.96 |
| 419s | Jiang_Server | 57 | -41.96 | -43.29 | -43.55 | -41.24 | -42.51 | -0.75 | -0.75 | -32.45 |
| 420 | Handl | 30 | -33.34 | -34.78 | -51.04 | -58.61 | -44.44 | -0.78 | -1.48 | -60.00 |
| 247s | BhageerathH | 57 | -42.07 | -38.75 | -55.31 | -49.25 | -46.35 | -0.81 | -0.81 | 49.09 |
| 201 | TsaiLab | 54 | -46.42 | -44.73 | -51.15 | -49.88 | -48.05 | -0.84 | -0.89 | -32.01 |
| 258s | Jiang_Threader | 57 | -45.72 | -46.32 | -56.04 | -52.68 | -50.19 | -0.88 | -0.88 | -33.20 |
| 474 | WAC_LABS | 57 | -49.06 | -45.91 | -59.39 | -53.72 | -52.02 | -0.91 | -0.91 | 2.99 |
| 111 | SIAT1068 | 56 | -50.71 | -46.77 | -64.70 | -55.56 | -54.44 | -0.96 | -0.97 | -27.24 |
| 259 | ALAdeGAP | 55 | -53.60 | -50.16 | -66.37 | -61.26 | -57.85 | -1.02 | -1.05 | -13.64 |
| 254 | CaspIta | 56 | -55.87 | -49.02 | -69.13 | -65.32 | -59.84 | -1.05 | -1.07 | 9.57 |
| 433 | Sun_Tsinghua | 41 | -60.28 | -55.52 | -57.31 | -67.87 | -60.25 | -1.06 | -1.47 | -66.46 |
| 195s | RBO-MBS | 56 | -63.23 | -59.53 | -59.89 | -63.25 | -61.48 | -1.08 | -1.10 | 135.54 |
| 492s | RBO-i-MBS | 57 | -65.07 | -61.94 | -60.89 | -64.27 | -63.04 | -1.11 | -1.11 | 138.14 |
| 117 | MBBS | 56 | -61.69 | -57.87 | -72.58 | -67.60 | -64.94 | -1.14 | -1.16 | -19.33 |
| 190s | RBO-MBS-BB | 57 | -67.72 | -64.70 | -64.85 | -67.05 | -66.08 | -1.16 | -1.16 | 136.17 |
| 107s | RBO-i-MBS-BB | 57 | -68.69 | -65.24 | -66.63 | -69.32 | -67.47 | -1.18 | -1.18 | 135.16 |
| 376 | Litonghua | 56 | -70.94 | -64.92 | -77.62 | -85.17 | -74.66 | -1.31 | -1.33 | -11.54 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 57 hAUs.

Supplementary Table S3. Sum and Average Z-scores for All Predictor Groups – 15 TBM_Hard Targets.

| Group | Name | N_model | GDT-HA | GDC-all | RPF | LDDT | Sum | Avg_4a | Avg_4s | MolPro |
|---|---|---|---|---|---|---|---|---|---|---|
| 489 | MULTICOM | 15 | 7.24 | 6.58 | 14.76 | 11.99 | 10.14 | 0.68 | 0.68 | 1.48 |
| 237 | zhang | 15 | 6.84 | 4.95 | 12.46 | 10.98 | 8.81 | 0.59 | 0.59 | -0.84 |
| 197 | Mufold | 14 | 6.02 | 3.92 | 12.09 | 10.40 | 8.11 | 0.54 | 0.58 | -6.17 |
| 267 | Pcons | 15 | 5.28 | 4.39 | 11.37 | 10.37 | 7.85 | 0.52 | 0.52 | 1.74 |
| 428 | PconsQ | 15 | 4.91 | 4.04 | 10.90 | 9.80 | 7.41 | 0.49 | 0.49 | -0.36 |
| 365 | chuo-fams | 15 | 4.68 | 2.88 | 11.77 | 10.17 | 7.38 | 0.49 | 0.49 | 1.99 |
| 27 | LEEcon | 15 | 5.08 | 3.38 | 10.55 | 9.44 | 7.11 | 0.47 | 0.47 | 3.17 |
| 130 | Pcomb | 15 | 4.55 | 3.05 | 11.10 | 9.17 | 6.97 | 0.46 | 0.46 | 8.60 |
| 035s | Zhang-Server | 15 | 5.18 | 3.55 | 9.85 | 8.88 | 6.87 | 0.46 | 0.46 | -3.49 |
| 344 | Jones-UCL | 14 | 4.30 | 3.47 | 10.25 | 8.54 | 6.64 | 0.44 | 0.47 | -13.90 |
| 294 | chuo-repack | 15 | 3.43 | 2.44 | 10.62 | 9.40 | 6.47 | 0.43 | 0.43 | -3.02 |
| 350 | Kloczkowski_Lab | 15 | 3.18 | 2.37 | 10.84 | 9.00 | 6.35 | 0.42 | 0.42 | -1.04 |
| 285 | McGuffin | 14 | 4.13 | 3.15 | 9.13 | 7.77 | 6.05 | 0.40 | 0.43 | -0.45 |
| 475 | CNIO | 15 | 3.55 | 1.79 | 9.99 | 8.62 | 5.99 | 0.40 | 0.40 | -1.58 |
| 79 | TASSER | 15 | 5.56 | 6.94 | 7.17 | 3.98 | 5.91 | 0.39 | 0.39 | -2.92 |
| 434 | chuo-fams-consensus | 15 | 2.83 | 1.55 | 9.51 | 8.67 | 5.64 | 0.38 | 0.38 | -0.49 |
| 114s | QUARK | 14 | 3.83 | 2.73 | 8.45 | 7.31 | 5.58 | 0.37 | 0.40 | -5.13 |
| 315 | keasar | 15 | 3.88 | 2.24 | 9.02 | 6.39 | 5.38 | 0.36 | 0.36 | -15.88 |
| 122s | RaptorX-ZY | 15 | 5.88 | 5.62 | 5.56 | 3.75 | 5.20 | 0.35 | 0.35 | -14.48 |
| 45 | Zhang_Ab_Initio | 15 | 2.96 | 3.37 | 7.68 | 6.71 | 5.18 | 0.35 | 0.35 | 1.19 |
| 490 | Zhang_Refinement | 15 | 4.00 | 2.56 | 7.52 | 6.40 | 5.12 | 0.34 | 0.34 | 4.24 |
| 26 | ProQ2clust | 14 | 2.78 | 3.41 | 6.79 | 6.68 | 4.92 | 0.33 | 0.35 | -1.75 |
| 101 | WeFold | 14 | 2.94 | 2.38 | 8.16 | 5.63 | 4.78 | 0.32 | 0.34 | -9.59 |
| 388 | ProQ2 | 15 | 2.11 | 1.11 | 8.25 | 7.48 | 4.74 | 0.32 | 0.32 | 20.29 |
| 330s | BAKER-ROSETTASERVER | 15 | 2.63 | 2.57 | 7.26 | 5.78 | 4.56 | 0.30 | 0.30 | 29.53 |
| 481 | Chicken_George | 15 | 2.44 | 2.05 | 7.19 | 5.24 | 4.23 | 0.28 | 0.28 | 1.41 |
| 458 | Sternberg | 15 | 3.47 | 2.69 | 4.70 | 5.80 | 4.17 | 0.28 | 0.28 | -6.16 |
| 477 | BAKER | 15 | 4.55 | 2.13 | 5.47 | 4.47 | 4.16 | 0.28 | 0.28 | 24.13 |
| 108s | PMS | 15 | 2.00 | 1.57 | 5.02 | 6.08 | 3.67 | 0.25 | 0.25 | -1.76 |
| 335s | TASSER-VMT | 15 | 4.24 | 4.06 | 3.55 | 2.50 | 3.59 | 0.24 | 0.24 | -10.74 |
| 479 | Boniecki_LoCoGRef | 11 | 1.24 | 1.76 | 5.49 | 5.07 | 3.39 | 0.23 | 0.31 | -6.66 |
| 405 | Mufold2 | 14 | 2.52 | 2.15 | 5.07 | 3.69 | 3.36 | 0.22 | 0.24 | -3.98 |
| 125s | MULTICOM-REFINE | 15 | 2.08 | 2.39 | 4.61 | 3.88 | 3.24 | 0.22 | 0.22 | 1.82 |
| 301 | LEE | 15 | 1.07 | 2.60 | 4.00 | 4.80 | 3.12 | 0.21 | 0.21 | -0.80 |
| 281 | Ariadne | 15 | 1.77 | 0.19 | 5.34 | 4.72 | 3.01 | 0.20 | 0.20 | 8.41 |
| 486s | RaptorX | 15 | 5.53 | 6.33 | 0.23 | -1.48 | 2.65 | 0.18 | 0.18 | -5.86 |
| 178 | Bilab | 15 | 1.40 | 1.34 | 3.31 | 4.51 | 2.64 | 0.18 | 0.18 | -0.79 |
| 103s | PconsM | 15 | 0.90 | 0.65 | 5.22 | 3.75 | 2.63 | 0.18 | 0.18 | 5.16 |
| 113s | SAM-T08-server | 12 | 2.40 | 1.65 | 3.45 | 2.50 | 2.50 | 0.17 | 0.21 | 2.35 |
| 298 | MidwayFolding | 12 | 5.28 | 4.16 | 1.91 | -1.31 | 2.51 | 0.17 | 0.21 | -13.66 |
| 424s | MULTICOM-NOVEL | 15 | 1.20 | 0.48 | 4.38 | 2.98 | 2.26 | 0.15 | 0.15 | 3.53 |
| 164 | 4_BODY_POTENTIALS | 15 | -0.41 | -0.48 | 5.93 | 3.12 | 2.04 | 0.14 | 0.14 | 0.99 |
| 222s | MULTICOM-CONSTRUCT | 15 | 0.14 | 0.34 | 3.00 | 3.51 | 1.75 | 0.12 | 0.12 | 0.07 |
| 435 | ossia | 15 | 1.37 | 0.78 | 2.82 | 1.92 | 1.72 | 0.12 | 0.12 | -8.75 |
| 370s | HHpred-thread | 14 | 1.05 | 1.44 | 1.29 | 3.00 | 1.70 | 0.11 | 0.12 | -20.61 |
| 29 | shisen | 12 | 0.07 | 1.23 | 1.89 | 2.23 | 1.36 | 0.09 | 0.11 | 16.84 |
| 141 | Bates_BMM | 14 | 0.64 | -0.43 | 3.05 | 1.42 | 1.17 | 0.08 | 0.08 | -5.80 |
| 317 | SHORTLE | 11 | 0.57 | 0.22 | 3.01 | 0.61 | 1.10 | 0.07 | 0.10 | 6.48 |
| 453 | KnowMIN | 15 | -0.21 | 0.54 | 3.10 | 0.88 | 1.08 | 0.07 | 0.07 | 14.72 |
| 081s | MULTICOM-CLUSTER | 15 | -0.30 | -0.89 | 2.80 | 2.47 | 1.02 | 0.07 | 0.07 | 2.70 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 15 TBM_hard AUs.

Supplementary Table S3. Sum and Average Z-scores for All Predictor Groups – 15 TBM_Hard Targets.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 223s | HHpredAQ | 15 | 0.40 | 0.48 | 0.48 | 1.64 | 0.75 | 0.05 | 0.05 | -19.63 |
| 430s | HHpredA | 15 | 0.41 | 0.47 | 0.48 | 1.64 | 0.75 | 0.05 | 0.05 | -19.66 |
| 280 | ProQ2clust2 | 15 | -0.07 | -1.08 | 1.85 | 2.06 | 0.69 | 0.05 | 0.05 | 1.16 |
| 215 | ppfld | 2 | 0.29 | 0.86 | 0.41 | 0.61 | 0.54 | 0.04 | 0.27 | 0.00 |
| 292s | Pcons-net | 15 | -0.51 | 0.34 | 2.44 | -0.32 | 0.49 | 0.03 | 0.03 | 8.22 |
| 493 | LEEMO | 15 | -0.16 | -1.36 | 1.56 | 1.67 | 0.43 | 0.03 | 0.03 | -6.54 |
| 85 | Anthropic_Dreams | 2 | -0.22 | -0.05 | 1.05 | 0.54 | 0.33 | 0.02 | 0.17 | 3.87 |
| 149 | wfFUIK | 2 | 0.33 | 0.20 | 0.33 | 0.38 | 0.31 | 0.02 | 0.16 | 2.57 |
| 165 | Void_Crushers | 2 | -0.08 | -0.10 | 0.76 | 0.71 | 0.32 | 0.02 | 0.16 | 4.37 |
| 68 | FOLDIT | 2 | -0.45 | 0.11 | 0.42 | 0.21 | 0.07 | 0.01 | 0.04 | 3.88 |
| 341 | Contenders | 1 | -0.27 | -0.43 | 0.35 | 0.03 | -0.08 | -0.01 | -0.08 | 1.70 |
| 131 | BioNanopore | 1 | -0.24 | -0.20 | 0.06 | -0.05 | -0.11 | -0.01 | -0.11 | 0.43 |
| 260 | wfFUGT | 1 | -0.07 | -0.45 | -0.24 | 0.03 | -0.18 | -0.01 | -0.18 | -0.74 |
| 311 | Laufer | 1 | -0.27 | -0.44 | -0.05 | 0.03 | -0.18 | -0.01 | -0.18 | 1.33 |
| 488s | chunk-TASSER | 15 | 0.63 | -0.48 | -0.98 | -0.06 | -0.22 | -0.02 | -0.02 | -6.49 |
| 251 | laufercenter_meta | 1 | -0.03 | -0.31 | -0.44 | -0.27 | -0.26 | -0.02 | -0.26 | 2.62 |
| 356s | sysimm | 2 | -0.01 | 0.08 | -0.39 | -1.58 | -0.48 | -0.03 | -0.24 | 3.99 |
| 448 | KIAS-Gdansk | 2 | -0.96 | -0.83 | -0.43 | 0.00 | -0.56 | -0.04 | -0.28 | -2.95 |
| 172 | Zhang-IRU | 13 | -2.15 | -1.04 | 1.04 | -0.64 | -0.70 | -0.05 | -0.05 | 32.67 |
| 77 | FLOUDAS | 10 | -0.91 | -2.57 | 0.65 | -0.27 | -0.78 | -0.05 | -0.08 | 6.16 |
| 287 | wfCPUNK | 3 | -1.12 | -0.93 | -0.67 | -0.49 | -0.80 | -0.05 | -0.27 | -1.44 |
| 411s | FALCON-TOPO | 15 | -0.77 | -0.29 | -0.02 | -2.14 | -0.81 | -0.05 | -0.05 | -4.51 |
| 456s | FALCON-TOPO-X | 15 | -0.87 | -0.14 | -0.40 | -2.85 | -1.07 | -0.07 | -0.07 | -4.60 |
| 373 | Kim_Kihara | 15 | -1.87 | -2.85 | -0.02 | -0.27 | -1.25 | -0.08 | -0.08 | 13.78 |
| 441 | WeFoldMix | 1 | -1.77 | -1.45 | -0.84 | -1.04 | -1.28 | -0.09 | -1.28 | 2.09 |
| 028s | YASARA | 3 | -1.26 | -1.72 | -1.78 | -0.99 | -1.44 | -0.10 | -0.48 | 6.10 |
| 286s | Mufold-MD | 15 | -1.48 | -0.87 | -1.43 | -2.43 | -1.55 | -0.10 | -0.10 | 8.61 |
| 124s | PconsD | 15 | -2.64 | -2.38 | 0.29 | -1.61 | -1.59 | -0.11 | -0.11 | 0.80 |
| 121s | STRINGS | 12 | -2.95 | -2.15 | -1.25 | -1.04 | -1.85 | -0.12 | -0.15 | -2.19 |
| 437 | sumalab | 15 | -0.32 | -2.17 | -2.82 | -2.46 | -1.94 | -0.13 | -0.13 | -2.88 |
| 148s | MUFold_CRF | 14 | -1.33 | -1.12 | -2.34 | -3.09 | -1.97 | -0.13 | -0.14 | -9.20 |
| 155 | Graham | 2 | -1.56 | -1.20 | -2.35 | -3.22 | -2.08 | -0.14 | -1.04 | -3.96 |
| 413s | ZHOU-SPARKS-X | 15 | -3.06 | -1.91 | -2.43 | -0.97 | -2.09 | -0.14 | -0.14 | -10.03 |
| 098s | confuzzGS | 3 | -2.26 | -1.55 | -2.96 | -2.00 | -2.19 | -0.15 | -0.73 | -2.36 |
| 087s | Distill_roll | 15 | -3.70 | -2.81 | 0.29 | -2.63 | -2.21 | -0.15 | -0.15 | 7.63 |
| 302s | 3D-JIGSAW_V5-0 | 14 | -0.36 | -0.94 | -2.53 | -5.37 | -2.30 | -0.15 | -0.16 | -2.82 |
| 43 | sessions | 15 | -4.03 | -4.16 | -0.69 | -0.74 | -2.41 | -0.16 | -0.16 | -8.91 |
| 444 | Lenregular | 2 | -1.48 | -2.06 | -2.79 | -4.00 | -2.58 | -0.17 | -1.29 | -4.00 |
| 152 | Cornell-Gdansk | 8 | -4.11 | -4.24 | -0.61 | -1.50 | -2.62 | -0.17 | -0.33 | -12.17 |
| 358s | RaptorX-Roll | 7 | -2.62 | -2.71 | -2.71 | -2.85 | -2.72 | -0.18 | -0.39 | -1.05 |
| 51 | thyzju | 2 | -2.79 | -2.52 | -2.83 | -3.17 | -2.83 | -0.19 | -1.41 | -0.75 |
| 179s | Lenserver | 3 | -0.44 | -0.87 | -4.78 | -6.00 | -3.02 | -0.20 | -1.01 | -6.00 |
| 348s | Phyre2_A | 15 | -3.73 | -3.08 | -3.78 | -1.65 | -3.06 | -0.20 | -0.20 | -14.75 |
| 494s | GSmetaserver | 10 | -1.21 | -1.11 | -3.70 | -6.37 | -3.10 | -0.21 | -0.31 | -6.10 |
| 333s | MUFOLD-Server | 15 | -1.91 | -1.47 | -4.72 | -4.80 | -3.23 | -0.22 | -0.22 | -3.28 |
| 462s | confuzz3d | 3 | -3.28 | -3.01 | -4.02 | -4.05 | -3.59 | -0.24 | -1.20 | -5.43 |
| 473 | Seok | 15 | -3.44 | -2.49 | -5.12 | -3.40 | -3.61 | -0.24 | -0.24 | 8.29 |
| 277s | Bilab-ENABLE | 15 | -4.07 | -3.90 | -3.12 | -3.48 | -3.64 | -0.24 | -0.24 | 4.30 |
| 175s | FRESS_server | 14 | -5.97 | -4.83 | -4.73 | -0.26 | -3.95 | -0.26 | -0.28 | -22.68 |
| 300 | PAX | 13 | -2.98 | -2.39 | -5.89 | -4.71 | -3.99 | -0.27 | -0.31 | 10.77 |
| 343s | NewSerf | 15 | -3.67 | -2.81 | -5.53 | -4.28 | -4.07 | -0.27 | -0.27 | -12.74 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 15 TBM_hard AUs.

Supplementary Table S3. Sum and Average Z-scores for All Predictor Groups – 15 TBM_Hard Targets.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Taylor | 5 | -3.44 | -3.69 | -5.51 | -4.17 | -4.20 | -0.28 | -0.84 | -7.80 |
| MeilerLab | 6 | -4.56 | -4.00 | -5.55 | -4.14 | -4.56 | -0.30 | -0.76 | 1.60 |
| Jiang_Human | 15 | -4.71 | -4.51 | -4.60 | -4.43 | -4.56 | -0.30 | -0.30 | -7.09 |
| samcha-server | 13 | -5.06 | -3.02 | -4.64 | -6.33 | -4.76 | -0.32 | -0.37 | -2.72 |
| Atome2_CBS | 11 | -4.39 | -2.18 | -5.09 | -8.09 | -4.94 | -0.33 | -0.45 | -5.69 |
| IntFOLD2 | 15 | -5.22 | -4.54 | -5.57 | -5.08 | -5.10 | -0.34 | -0.34 | -8.85 |
| Distill | 15 | -5.00 | -4.50 | -5.20 | -6.14 | -5.21 | -0.35 | -0.35 | 1.38 |
| Jiang_Server | 15 | -4.71 | -4.61 | -6.90 | -4.77 | -5.25 | -0.35 | -0.35 | -7.17 |
| hGen3D | 15 | -5.13 | -5.73 | -5.16 | -5.17 | -5.30 | -0.35 | -0.35 | -9.16 |
| chuo-fams-server | 14 | -5.80 | -5.99 | -5.44 | -4.75 | -5.50 | -0.37 | -0.39 | -11.13 |
| PROTAGORAS | 6 | -4.73 | -4.73 | -7.05 | -6.78 | -5.82 | -0.39 | -0.97 | -1.59 |
| Seok-server | 15 | -3.46 | -3.29 | -10.22 | -6.89 | -5.97 | -0.40 | -0.40 | 7.08 |
| chuo-repack-server | 15 | -5.78 | -5.90 | -6.72 | -7.21 | -6.40 | -0.43 | -0.43 | -12.04 |
| POEMhome | 9 | -6.56 | -5.05 | -7.20 | -7.47 | -6.57 | -0.44 | -0.73 | -3.49 |
| slbio | 14 | -4.76 | -2.40 | -7.60 | -12.03 | -6.70 | -0.45 | -0.48 | 8.17 |
| SAM-T06-server | 13 | -6.93 | -5.80 | -7.54 | -6.65 | -6.73 | -0.45 | -0.52 | 5.03 |
| UGACSBL | 13 | -6.75 | -5.70 | -9.04 | -6.05 | -6.89 | -0.46 | -0.53 | -6.39 |
| SIAT1068 | 14 | -5.50 | -5.07 | -10.48 | -7.25 | -7.08 | -0.47 | -0.51 | -4.91 |
| TsaiLab | 12 | -5.67 | -5.47 | -9.61 | -8.04 | -7.20 | -0.48 | -0.60 | -5.45 |
| BhageerathH | 15 | -5.62 | -5.38 | -9.86 | -8.37 | -7.31 | -0.49 | -0.49 | 14.12 |
| MATRIX | 13 | -5.97 | -5.70 | -9.07 | -8.76 | -7.38 | -0.49 | -0.57 | 7.73 |
| FFAS03mt | 10 | -3.53 | -2.74 | -12.06 | -12.64 | -7.74 | -0.52 | -0.77 | -1.83 |
| FFAS03c | 15 | -5.47 | -3.53 | -12.31 | -11.69 | -8.25 | -0.55 | -0.55 | -8.70 |
| RBO-MBS | 14 | -8.58 | -7.84 | -7.80 | -8.89 | -8.28 | -0.55 | -0.59 | 34.22 |
| IntFOLD | 15 | -6.41 | -6.37 | -10.05 | -10.59 | -8.36 | -0.56 | -0.56 | -2.26 |
| FFAS03 | 11 | -4.36 | -3.13 | -12.80 | -13.38 | -8.42 | -0.56 | -0.77 | -3.34 |
| MicroSumaLab | 8 | -7.13 | -7.23 | -10.00 | -9.55 | -8.48 | -0.57 | -1.06 | 7.49 |
| panther | 9 | -7.18 | -4.93 | -10.60 | -11.53 | -8.56 | -0.57 | -0.95 | -4.49 |
| Jiang_Fold | 15 | -7.66 | -6.86 | -11.33 | -11.10 | -9.24 | -0.62 | -0.62 | -6.43 |
| HandI | 8 | -4.37 | -5.08 | -12.40 | -16.00 | -9.46 | -0.63 | -1.18 | -16.00 |
| FFAS03hj | 13 | -7.63 | -5.76 | -11.21 | -14.29 | -9.72 | -0.65 | -0.75 | -3.05 |
| RBO-i-MBS-BB | 15 | -9.97 | -9.47 | -8.93 | -11.65 | -10.01 | -0.67 | -0.67 | 35.03 |
| AOBA-server | 14 | -6.12 | -4.90 | -13.09 | -16.29 | -10.10 | -0.67 | -0.72 | -3.87 |
| RBO-i-MBS | 15 | -9.95 | -9.61 | -9.82 | -11.07 | -10.11 | -0.67 | -0.67 | 36.05 |
| HOMER | 11 | -5.41 | -5.26 | -13.54 | -17.03 | -10.31 | -0.69 | -0.94 | 12.28 |
| chuo-binding-sites | 15 | -9.46 | -7.61 | -12.81 | -12.43 | -10.58 | -0.71 | -0.71 | -13.63 |
| Sun_Tsinghua | 10 | -9.98 | -8.95 | -9.71 | -14.00 | -10.66 | -0.71 | -1.07 | -15.90 |
| RBO-MBS-BB | 15 | -10.52 | -11.04 | -9.42 | -11.81 | -10.70 | -0.71 | -0.71 | 35.35 |
| biouv | 9 | -10.04 | -9.54 | -11.00 | -12.31 | -10.72 | -0.72 | -1.19 | -14.94 |
| Jiang_Threader | 15 | -7.94 | -7.60 | -15.42 | -12.16 | -10.78 | -0.72 | -0.72 | -6.74 |
| CaspIta | 15 | -9.27 | -7.05 | -14.62 | -14.08 | -11.26 | -0.75 | -0.75 | -0.34 |
| WAC_LABS | 15 | -11.24 | -10.05 | -15.83 | -13.02 | -12.54 | -0.84 | -0.84 | -1.77 |
| MBBS | 15 | -11.57 | -9.69 | -16.13 | -13.93 | -12.83 | -0.86 | -0.86 | -3.82 |
| ALAdeGAP | 14 | -11.14 | -9.15 | -16.96 | -15.06 | -13.08 | -0.87 | -0.93 | -2.81 |
| Litonghua | 15 | -11.23 | -9.34 | -14.32 | -17.76 | -13.16 | -0.88 | -0.88 | -1.75 |



The columns labeled GDT-HA, GDC-all, RPF, and LDDT are the sum of Z scores across all models submitted by each predictor group. The Sum column is the average of the sum of Z scores for the four metrics assessed. The Avg-4s and Avg-4a scores are the Sum scored divided by the number of AUs for which a model was submitted (s) by each predictor group, and the total number of AUs used for assessment (a), respectively. These scores are identical for predictor groups who submitted models for all 15 TBM_hard AUs.

## Supplementary Figures

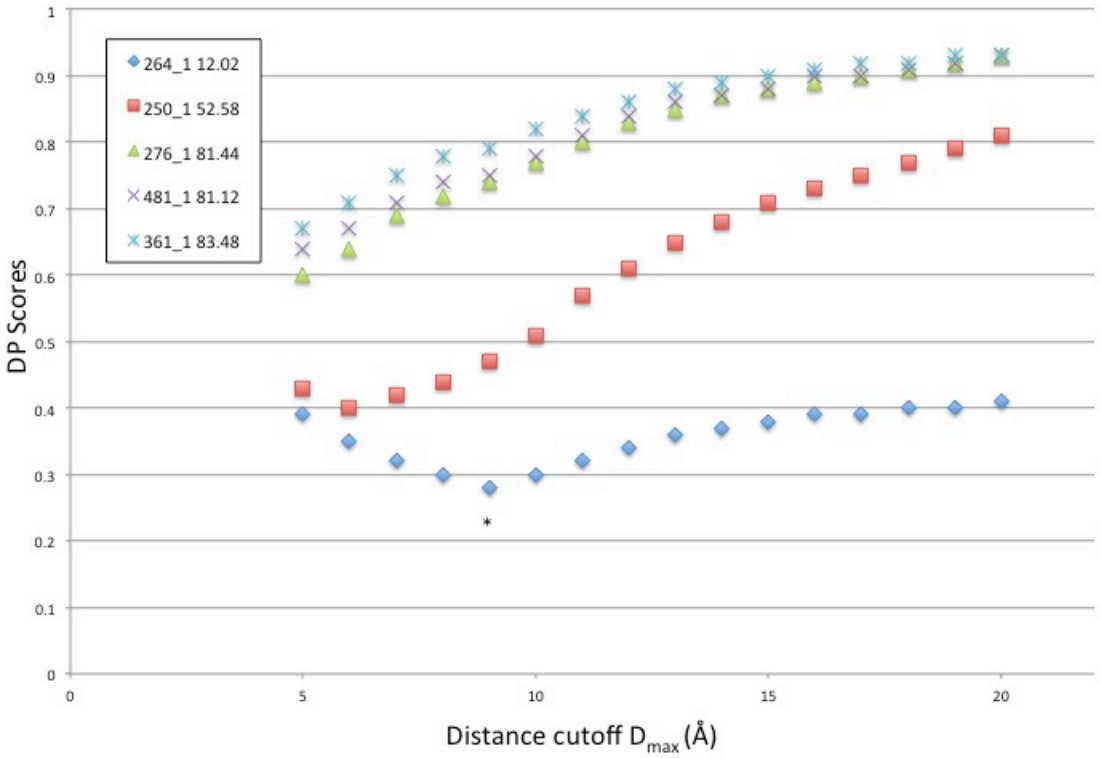


**Figure S1. Selection of $D_{max}$.** DP scores of a CASP9 target T0570 for values of $D_{max}$ ranging from 5 to 20 Å are shown for models 264_1, 250_1, 276_1, 481_1 and 361_1. The corresponding GDT-TS scores for each model are also shown as inset in the graph. The best discrimination between these models is obtained with $D_{max}$ = 9.0 Å (indicated by the $D_{max}$ value labeled *). This $D_{max}$ shows optimum discrimination for models in several accuracy ranges, including among the models with similar GDT_TS scores ≥ 80 (e.g. among models 361_1, 481_1, 276_1), models with 40 < GDT_TS < 80 (e.g. model 250_1) and models with GTD_TS ≤ 40 groups (e.g. model 264_1).

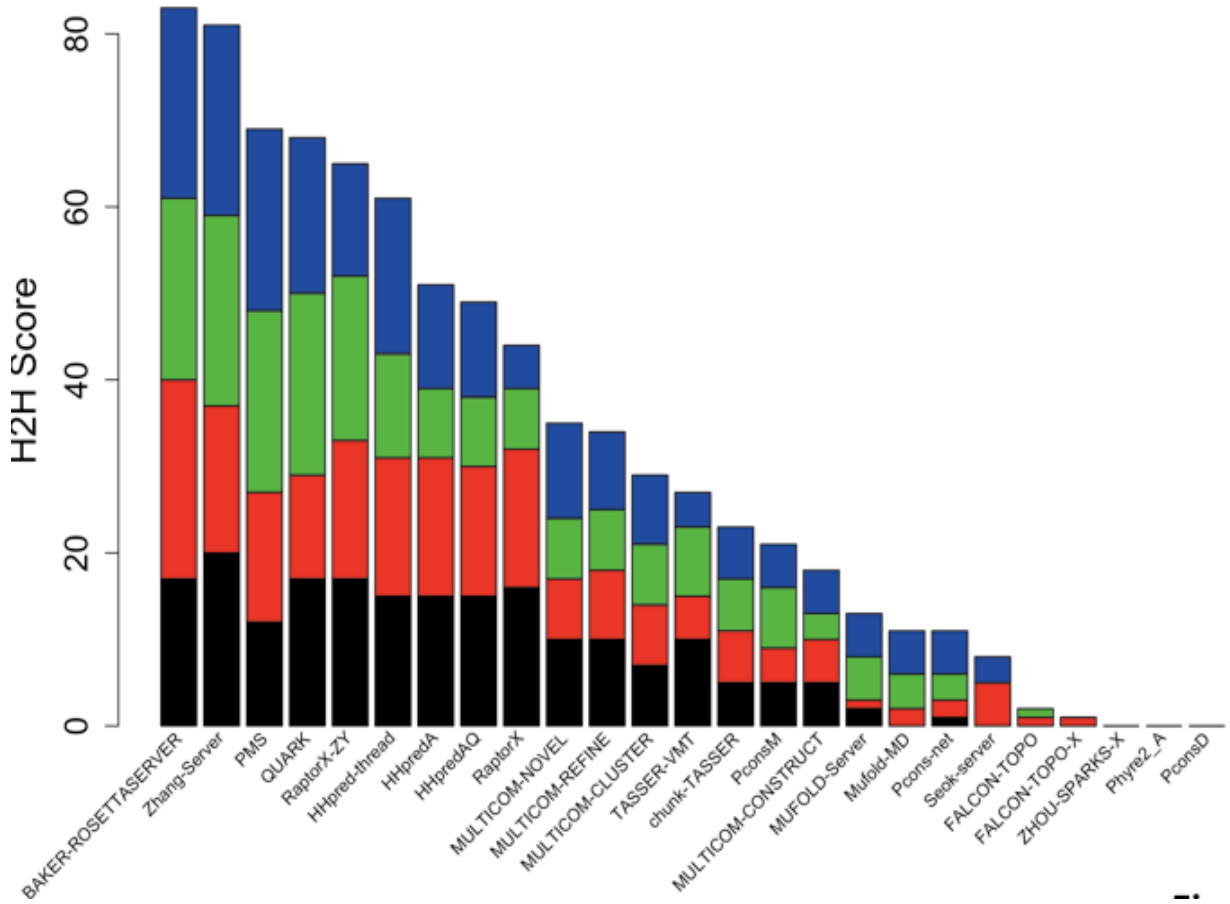


**Figure S2**. **Head-to-head pairwise Wilcoxon rank sum analysis on raw scores between 25 top-ranking server competitor groups for 112 hsAUs**. We carried out the Wilcoxon signed rank test of the two paired samples using the same dataset used for the rankings shown in Fig. 4A. Ties receive a rank equal to the average of the ranks they span, and normal approximation was used to calculate the p-value because the number of samples is larger than 50. Black, GDT-HA; red, GDC-all; green, RPF-9; blue, LDDT-15. Top-ranking 25 groups were identified based on average Z score (Table I).

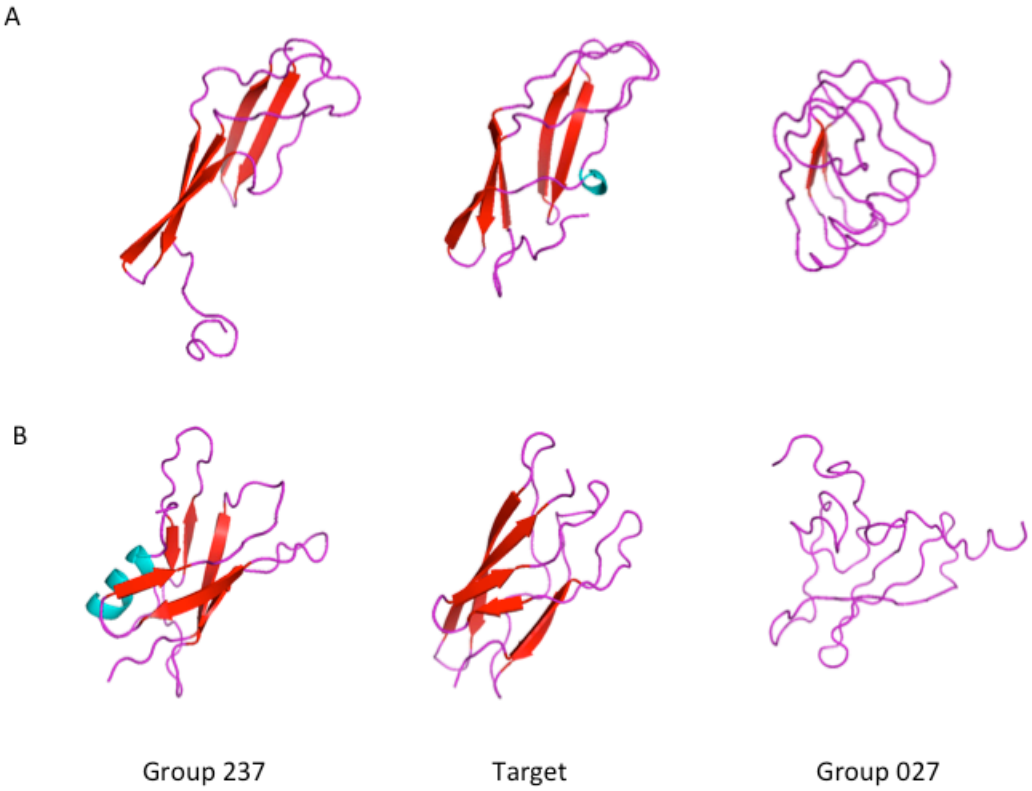


**Figure S3**. Ribbon diagrams of models 237_1 (left), 027_1 (right) and the target experimental structures (middle) for CASP10 AU targets (A) T0671-D1 and (B) T0705-D1. Beta sheets are colored in red, helices are colored in cyan and loops are colored in magenta.

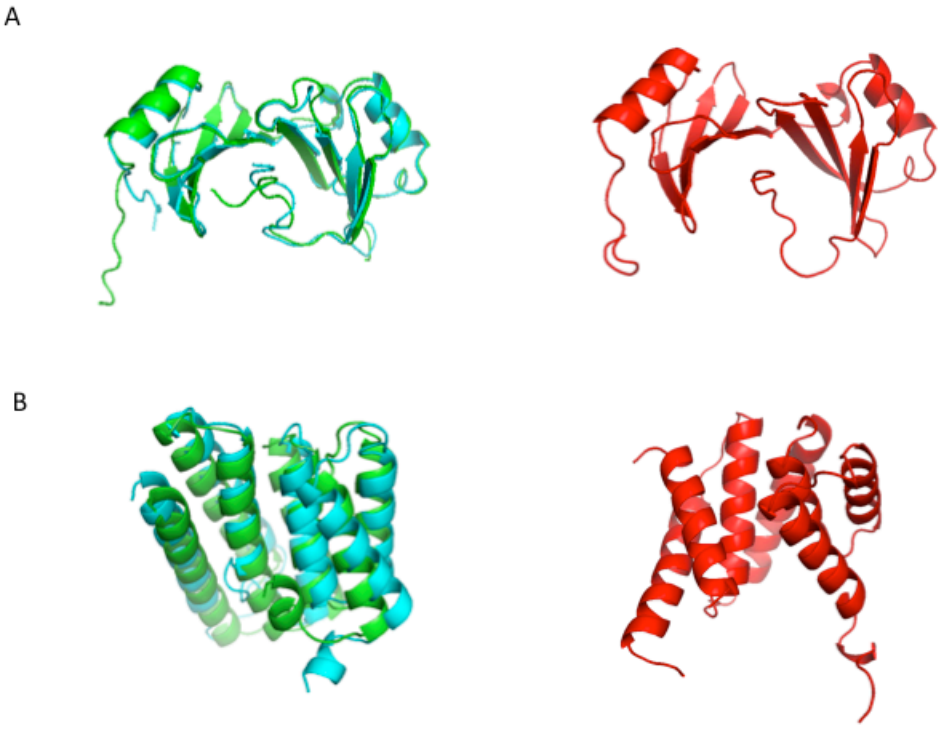


**Figure S4.** (A) Left: Superimposed ribbon diagrams of CASP10 prediction models 101_1(cyan) and 113_1(green) for target T0644-D1. Right: the target experimental structure (red) (B) Left: Superimposed ribbon diagram of CASP10 prediction models from 237_1 (green) and 079_1 (cyana) for AU target T0678-D1. Right: the target experimental structure (red).

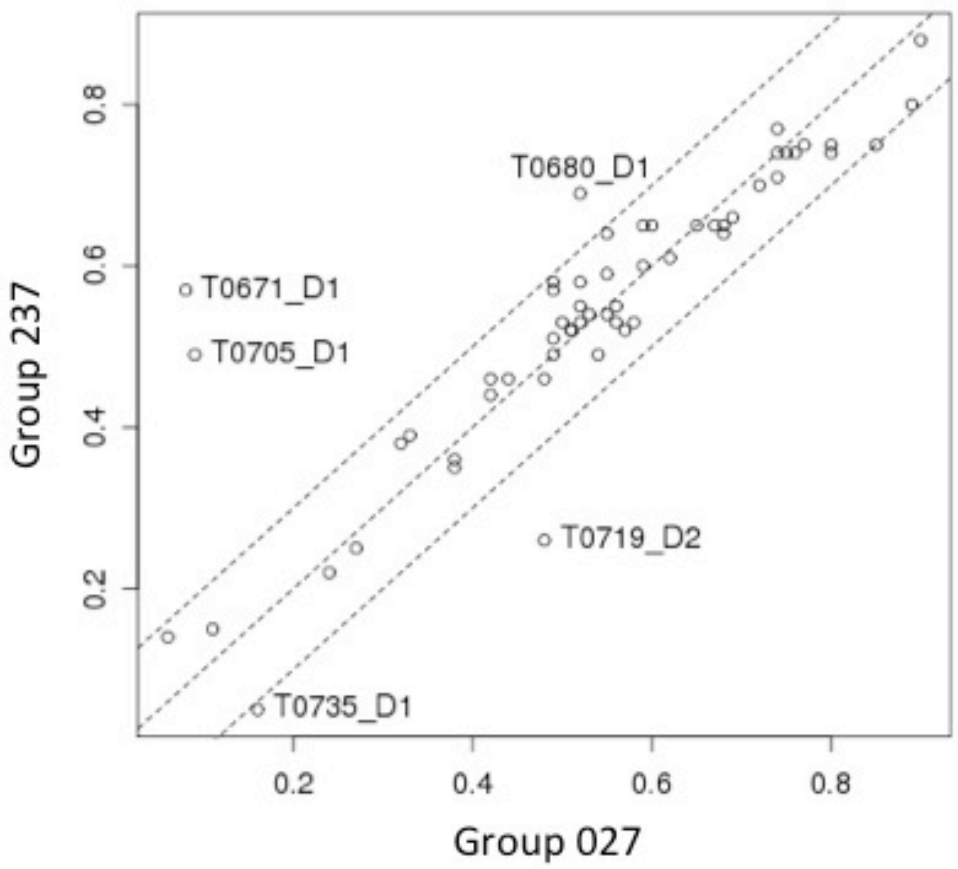


**Figure S5**. **Scatter plot of RPF DP raw scores for all the Human and/or Server AUs between groups 237 and 027**.

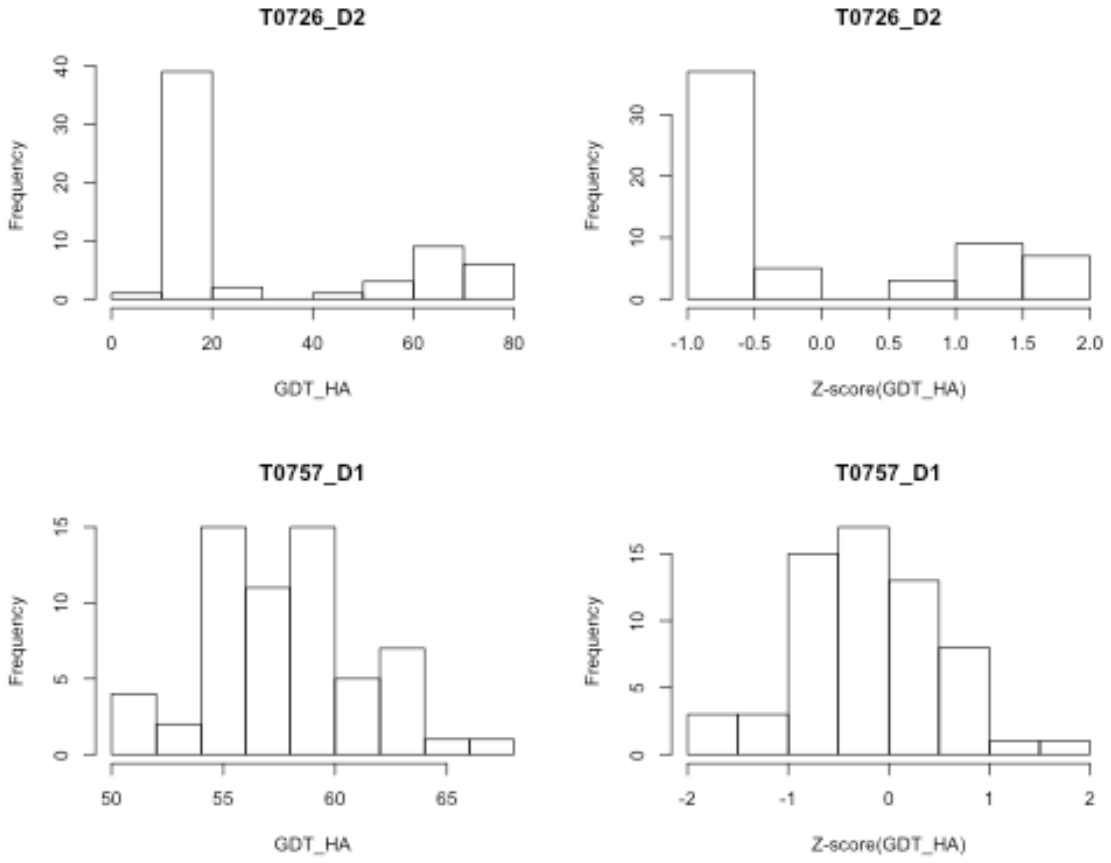


**Figure S6. Raw scores are more sensitive to outliers than Z scores.** We considered two commonly observed scenarios. In the first example, observed for AU T0726_D2 (top two panels), the predictions fall into two distinct groups (i.e., either Good or Bad predictions). Clearly, the distribution of the raw scores violates the normal assumption and the normalized Z-score is not appropriate for the head-to-head comparison. More specifically, for T0726_D2, one group has a prediction with GDT-HA=74.69, Z-score (GDT-HA)=1.76, and another group has a prediction with GDT-HA=29.40, Z-score (GDT-HA)=-0.04. Their raw score difference is 45.29 and Z-score difference is 1.8. The raw score difference is quite large, while the Z score difference is relatively small. The second scenario is observed for AUs including T0757_D1. Here, all the groups have similar predictions, therefore the spread of the raw scores is quite narrow, and Z-score tends to over-estimate the degree of divergence. For T0757_D1, one group has a prediction with GDT-HA = 67.11, Z-score(GDT-HA) = 1.58, and the

other group has a prediction with GDT-HA = 57.59, Z-score(GDT-HA) = -0.21. Their raw scores difference is 9.52 and Z-score difference is 1.79. This is about the same Z score difference as the predictions for AU T0726_D2, but with a much smaller raw score difference. These data demonstrate our common observation that raw scores can better discriminate certain scenarios than the Z-scores.

**Supplementary References**


1. Rosato A, Aramini JM, Arrowsmith C, Bagaria A, Baker D, Cavalli A, Doreleijers JF, Eletsky A, Giachetti A, Guerry P, Gutmanas A, Guntert P, He Y, Herrmann T, Huang YJ, Jaravine V, Jonker HR, Kennedy MA, Lange OF, Liu G, Malliavin TE, Mani R, Mao B, Montelione GT, Nilges M, Rossi P, van der Schot G, Schwalbe H, Szyperski TA, Vendruscolo M, Vernon R, Vranken WF, de Vries S, Vuister GW, Wu B, Yang Y, Bonvin AM. Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 2012;20(2):227-236.

2. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. Proteins 2011;79 Suppl 10:37-58.